Obstacles and opportunities in meta-analysis of genetic association studies

Georgia Salanti, PhD[1], Simon Sanderson, DPH[2], and Julian P.T. Higgins, PhD[3]

Genetic association studies have the potential to advance our understanding of genotype-phenotype relationships, especially for common, complex diseases where other approaches, such as linkage, are less powerful. Unfortunately, many reported studies are not replicated or corroborated. This lack of reproducibility has many potential causes, relating to study design, sample size, and power issues, and from sources of true variability among populations. Genetic association studies can be considered as more similar to randomized trials than other types of observational epidemiological studies because of "Mendelian randomization" (Mendel's second law). The rationale and methodology for synthesizing randomized trials is highly relevant to the meta-analysis of genetic association studies. Nevertheless, there are a number of obstacles to overcome when performing such meta-analyses. In this review, the impacts of Type I error, lack of power, and publication and reporting biases are explored, and the role of multiple testing is discussed. A number of special features of association studies are especially pertinent, because they may lead to true variability among study results. These include population dynamics and structure, linkage disequilibrium, conformity to Hardy-Weinberg Equilibrium, bias, population stratification, statistical heterogeneity, epistatic and environmental interactions, and the choice of statistical models used in the analysis. Approaches to dealing with these issues are outlined. The supreme importance of complete and consistent study reporting and of making data readily available is also highlighted as a prerequisite for sound meta-analysis. We believe that systematic review and meta-analysis has an important role to play in understanding genetic association studies and should help us to separate the wheat from the chaff. *Genet Med* 2005:7(1):13–20.

Key Words: *meta-analysis, heterogeneity, bias, genetic association studies, population characteristics*

Genetic association studies assess correlations between disease status and genetic variants in a population.[1] They have been particularly advocated for the investigation of complex, chronic diseases where other approaches (such as linkage) are less powerful. Advances in genotyping techniques and the discovery of a huge number of variants in the human genome have led to a proliferation of association studies, many of which have not led to advances in our understanding of disease.[2] Sadly, the literature contains many reported associations that cannot be replicated or supported by linkage or functional studies.[1,3] This lack of reproducibility stems from a number of causes, related to study design, sample size, and power issues, and true variability between populations.[4]

Meta-analysis provides an opportunity to help identify genuine associations by addressing some of these obstacles.[5,6] Re-cently, an eloquent introduction to meta-analysis as applied to genetic association studies has been published.[7] Although methods for meta-analysis of randomized trials are well developed, methods for observational studies lag some way behind.[8] However, some have argued that genetic association studies are more closely related to randomized trials than other types of epidemiological study because of "Mendelian randomization" (Mendel's second law).[9,10] The random allocation of alleles at any particular locus may provide a mechanism similar to randomization in clinical trials. This random assortment of alleles theoretically should be independent of environmental factors. The rationale and methodology for synthesizing randomized trials may therefore be highly relevant to the meta-analysis of genetic association studies.

Our aim in this review is to identify key issues and their implications for the meta-analysis of genetic association studies. The article is organized in four sections. In the first section, we discuss the role of chance (type I error), publication, and reporting biases. In the second section, we support these views using evidence from published studies on these topics. The third section previews the special features of genetic association studies that may lead to true variability in associations across studies. These include the impact of population structure and dynamics, conformity to Hardy-Weinberg Equilib-

rium, and the choice of statistical models used in the meta-analysis. Finally, we discuss how all of these issues can be addressed when performing meta-analysis. We conclude with a plea that the reporting of genetic association studies should be complete and consistent to facilitate sound meta-analysis.

CHANCE EFFECTS AND MULTIPLE GENES

Inconsistency of findings across studies can be due to false negatives (underpowered studies), false positives (spurious findings), or true variability across different populations.[11] Given a large and representative selection of studies of a particular association, the first two would not cause concern, as the false-negative and false-positive findings would cancel each other out across studies. However, in reality, the availability of findings may be expected to depend on the significance of the result as perceived by the investigators. Thus, false-negative results may be suppressed and false-positive results given prominence.

Lack of power

Genetic association studies have a tendency to lack the power to detect a statistically significant association. Indeed, the most realistic genetic association between a polymorphic locus and a disease has been claimed to yield an odds ratio between 1.1 and 1.5.[12] Therefore, to achieve a satisfactory power, at least 1000 subjects are required, and often more, depending on the prevalence of the polymorphism. Lohmueller et al.[11] highlighted this problem, observing that true effect magnitudes based on meta-analyses of multiple studies require sample sizes for individual studies of several thousands subjects. Whereas some large cohort studies have been undertaken recently, for many studies such a large sample size is an unrealistic goal. Meta-analysis of multiple studies clearly has a role in offering an analysis with the potential for higher power.[7]

Because power calculations are based on expected effect magnitudes, usually derived from previous studies, overoptimistic early findings can lead to subsequent studies being underpowered. This has been shown to be an important problem, especially when the true association is weak.[4,13]

Reporting biases and multiple comparisons

Lack of power is a long-standing problem for single studies, but can to an extent be addressed by the combination of multiple studies. However, the problem of false-positive studies has emerged as the key threat to the validity of meta-analyses of genetic association studies. Reporting biases is a well-established problem in meta-analysis but most attention has been given to the absence of entire studies on the basis of the statistical significance of the main finding. The inclusion of unpublished data as a means of reducing publication bias is commonly suggested although the ability to locate all unpublished studies is rare in practice.

A more important problem in meta-analysis of genetic association studies is bias due to within-study selective reporting. This occurs when multiple analyses have been performed but only a selected subset of them is reported. We can typically investigate many more potential genetic markers than environmental exposures, and testing multiple hypotheses about numerous loci or markers in a gene runs the risk of false-positive findings. Genetic association studies vary widely in the number of markers they investigate. Selection of markers is relatively straightforward if there is a variant of the candidate gene that is known to have a functional effect or is suspected to be directly responsible for predisposition to the disease. This should increase the prior probability of detecting a true association if one exists. Alternatively, genomic scans and comprehensive studies can cover very large numbers of markers.[1,14] Furthermore, quantitative traits may be affected by a large number of gene variants that only have a small impact on trait variability.[15]

Other candidates for multiple testing include multiple phenotypes (e.g., disease outcomes), multiple patient subgroups, and multiple statistical models (including different assumptions regarding mode of inheritance). Adjustments for multiple testing may temper the statistical significance of the findings, but may not be sufficient to prevent selective reporting bias. Constraints on the length of published articles can result in only the most exciting results being presented, even if the *P* values have been adjusted for multiple testing. Electronic publishing of extensive results tables and data depositories for studies of gene-disease association studies would go a long way toward making unbiased results available for meta-analysis.

A growing number of empirical studies are finding that early results relating to a particular marker may be overoptimistic, providing indirect evidence of selective reporting of exciting, yet spurious, findings. In the next section, we overview these empirical studies.

EMPIRICAL EVIDENCE OF LACK OF REPLICATION

Meta-analyses provide the strongest evidence when the findings of multiple well-conducted studies agree. A problem in genetic association studies that has attracted particular interest is agreement between the first published study and later results. This may occur primarily because of reporting bias, including simple publication bias, selective reporting, or differential time lag between conduct and reporting of the study according to the findings. These are not unusual observations in epidemiological research, but in the case of genetic association studies the problem is more important given the large number of potential exposures, many without strong rationale for being related to the outcome.

Ioannidis et al.[16] have examined a number of meta-analyses of genetic associations. Combining observations across meta-analyses, they found that the result of the first study showed a tendency to overestimate the true effect, particularly when the original study contained few participants. In addition, they observed that the greater the number of subsequent studies included in the meta-analysis, the larger the discrepancy between the pooled and first published result, although this may

at least partly be explained by increased power to detect the discrepancy.

The same authors addressed the problem of heterogeneity and bias in an extension of the above mentioned study.[17] They reviewed 55 meta-analyses of genetic associations, aiming to investigate whether the size of a study was associated with the outcome, to test whether any discrepancy between the results of the first and any subsequent trials appeared genuine (and determine the direction of any effect), to determine how often heterogeneity occurs, and finally, to test if and when these three situations coexist. The authors concluded that heterogeneity between studies occurred frequently and that the presence of heterogeneity was related to the presence of discrepancies either between small versus large studies or between the first versus subsequent studies. These two sources of discrepancy did not significantly coexist.

Using the same dataset of meta-analyses, Trikalinos et al.[18] revisited the problem of publication bias related to the first published study. They show that early published reports have minimal predictive value over the establishment of a significant association at the meta-analysis level. However, the conclusion of the first-published study has important implications for the pursuit of research on the given association: statistically significant first-published studies are followed by more studies over a longer period, whereas associations that are initially nonsignificant do not attract further research.

Hirschhorn et al.[13] scrutinized studies on 166 gene-disease associations. Only six collections of studies were considered to replicate findings. They concluded that early significant findings were often not reproduced in subsequent studies and argued that chance was an unlikely explanation of this finding. A combination of genuine sources of variation, such as different populations, the extent of linkage disequilibrium and gene-gene and gene-environment interactions was posed as a possible reason for the discrepant results. Furthermore, the authors suggested that low power to detect small genetic effects contributed to the lack of replication.

Further investigation of a subset of the same collections of studies on 25 gene-disease associations by Lohmueller et al.[11] addressed the problem of disagreement between the result of the first statistically significant study and the results of subsequent studies. They found that the association reported in the first study exceeded that from a meta-analysis of subsequent studies in 24 of the examples, highlighting the importance of not relying on initially statistically significant findings. However, in 11 of the meta-analyses there were at least two studies that replicated the first positive study, and eight of the meta-analyses showed a statistically significant association in agreement with the first positive report. The authors thus concluded that a sizable fraction of studies showed evidence of replication and that false-negative studies contributed to inconsistent replication. They argued that population admixture and publication bias were less likely explanations, and that meta-analysis is an essential tool until large, well-designed and well-reported studies become common.

SOURCES OF INCONSISTENCY

There are many reasons why gene-disease associations may genuinely vary among studies included in a meta-analysis. A key source of variation is the diversity in the populations studied, which may be an issue both within studies and across studies. Use of different phenotypic outcomes may also induce inconsistency. Variation in the methods used by the different studies leads to a different kind of inconsistency: variation in the extent of bias in the associations that the studies are evaluating. All of these sources of diversity might lead to variation in effects beyond that which may be expected due to chance alone, a situation commonly referred to as heterogeneity in the meta-analysis literature. Some of these issues have been discussed previously.[4] We review them briefly in this article, mainly from a hypothetical point of view: there is little empirical evidence of the relative importance of different potential sources of inconsistency.

Population characteristics

Linkage disequilibrium and population dynamics

A genetic marker that is targeted in an association study may not be a disease-causing locus itself, but may be linked to a causal one, such that passage of variants of the marker from one generation to the next are correlated with passage of variants of the disease-causing mutation. The resulting linkage disequilibrium (LD) will yield a significant association between the marker and the disease itself. The extent of LD between alleles reflects a population's recombination history. The further successive generations are from the original mutation the more recombinations occur, resulting in a smaller amount of shared DNA between individuals. LD can vary within and between populations because of regional variability of LD, genetic drift, population admixture, location chromosomal structure, and mating patterns.

These differences have been demonstrated between the UK and Finland. In Finland, the population has expanded from a relatively recent bottleneck, whereas in the UK the population has expanded gradually over many generations. Eaves investigated microsatellites from chromosome 18q21 in 664 British and 430 Finnish subjects: LD exists in both populations but extends over a longer range (up to 3 cM) in Finns.[19] Further, LD may occur in one population but not in another. For example, many of the observed associations with TNF-α may in fact reflect a true association with the HLA locus that has strong LD over large distances.[13]

True associations due to LD can show conflicting results depending on the population studied and the particular features of the DNA in those populations. Therefore, an attempt to assess LD may contribute to an understanding of why conflicting results arise in published studies and how best to deal with them in the context of meta-analysis. Whenever there is doubt about the nature of a disease-associated polymorphism, investigation of adjacent markers, say nearby SNPs in the same gene, may confirm whether the association is causal or due to LD. Characterization of background levels of LD, including

genomic features such as G+C levels, repetitive elements, and predicted or known hotspots may help in this regard.[20]

Allelic and locus heterogeneity

Both allelic heterogeneity (where many different mutations in the same gene can cause the same phenotype) and locus heterogeneity (where the same disease is caused by genetic variation in different loci in affected individuals) can result in different degrees of association among different studies. The association between any particular allele and disease, even if that allele has the same frequency in different populations, will depend on the frequency of any other disease-causing alleles. If these other alleles vary in frequency across populations, then different associations will be found. These variations are likely to have their own genetic ancestral background, different ancestral haplotypes, and nonrandom association patterns. An example of locus heterogeneity is provided by tuberous sclerosis, which is caused by mutations at either of two loci, TSC1 and TSC2. Familial hypercholesterolemia displays allelic heterogeneity, where over 700 disease-causing mutations of the low-density lipoprotein gene have been described to date. The frequency of these mutations varies extensively with the specific populations studied throughout the world.[21]

Isolated populations

Some genetic association studies are performed in populations that are more or less isolated. There are theoretical advantages to performing association studies in isolated populations such as Iceland or Finland, but they are often limited because the number of cases is too low to generate enough statistical power. Younger, isolated populations are more likely to show linkage disequilibrium. This could affect the presence and strength of any association detected. A degree of inbreeding can occur even in apparently mixed European populations (for example, within the Netherlands).[2]

Population stratification

Population stratification occurs when genetic variants are studied in samples that include a mixture of genetically distinct populations. If the studied disease is more common within a particular ethnic group, this group will be overrepresented in the cases. Therefore any polymorphism that marks genetically this ethnic group will appear to be associated with the outcome, clearly producing a false-positive finding. This effect is an important potential confounding variable and warrants careful thought and interpretation in assessing primary studies and their effects on meta-analysis.

For a confounding variable to be important it must have an effect of comparable magnitude to the main effect being investigated. The effect of population stratification has been assessed conceptually and empirically in certain populations and has been shown to be much less important than first thought.[22,23] However, the empirical evaluation was limited to the specific population studied (non-Hispanic Caucasians) and a single polymorphism (only *N*-acetyltransferase, NAT2), which reduces the applicability of the conclusions.[22] Edland[24]

also challenged the generalization of these findings and argued that ethnicity information should be used whenever available. On the other hand, Lohmueller et al.[11] exemplified how stratifying for ethnicity does not necessarily remove heterogeneity. Whereas it is widely acknowledged that situations where population stratification can be a significant confounder are rare, researchers do not advise that the issue should be ignored.[4,13] We agree that control for population stratification remains an important consideration, as it has the potential to expose genuine variation in the size of the association, including that caused by ethnicity-dependent penetrance and environmental differences.

Certain approaches in study design can mitigate the effect of population stratification. These methods include unlinked genome markers,[25] family-based design, and the use of parental controls and the transmission disequilibrium test.[13,26] Where this is not possible, each ethnic group can be analyzed separately, and association then tested within each group.[11]

Hardy-Weinberg equilibrium

Hardy-Weinberg equilibrium (HWE) refers to a situation in which the frequencies of genotypes are predicted by the frequencies of two alleles under a simple Mendelian inheritance model. The specific assumptions underlying HWE, including random mating, lack of selection according to genotype, and absence of mutation or migration, are rarely all met in human populations. However, population-based studies roughly approximate to HWE if they are large enough and they do not usually provide enough information to assess the size of departure from HWE.[27,28] Moreover, statistical tests for HWE are not powerful and can only detect large deviations, especially in small studies.

Genotyping error is an important cause of deviation from HWE.[28,29] Deviations may also occur in small populations due to genetic drift and founder effect, nonrandom mating (which occurs to some extent in nearly all groups but is especially common for some conditions such as deafness, epilepsy, and small stature), and heterozygote advantage (where heterozygotes have, or have had, some reproductive advantage over normal homozygotes; for example, in cystic fibrosis).

Confounding due to gene-gene and gene-environmental interaction

Confounding due to gene-gene epistatic interaction can occur when the studied trait is not caused by a gene alone, but by the effect of interactions between two or more loci. In addition, most common diseases arise as a result of an interactive mechanism between gene and environmental circumstances, which may vary both between and within populations. Thus interactions between the gene being studied and other factors are a potential explanation for inconsistency in results across studies. The lack of an association in a study does not necessarily exclude a gene-gene or gene-environment interactions, and thus the existence of a true association in a subgroup of participants.

Despite the problems of measurement of environmental factors, abandoning attempts to investigate these potential in-

teractions would be inappropriate.[30] Hirschorn et al.[13] argue that failure to address such confounding represents an important explanation for the nonreplication of genetic association studies. Little et al.[31] underline the need to include and accurately report environmental and genetic factors that potentially contribute to the manifestation of the disease in order to enhance their compatibility within a meta-analysis and reinforce the quality of the conclusions. As such, it is essential that both gene-environment and gene-gene interactions are considered at the level of both the individual studies, as well as in the meta-analysis.[4,13,32] Unfortunately, to overcome most confounding, one usually needs to perform a meta-analysis of individual participant data, which is not always feasible.[33]

Definition of phenotypes

Variation in the definition of clinical outcomes can be a major source of inconsistent results across studies. Some case-control studies may use extreme phenotypes. For example, studies of colon cancer cases can be restricted to those with extreme polyposis, a trait following a simple Mendelian pattern. For continuous traits, cases at the extremes of the distribution are often recruited to increase the possibility of detecting a genotype-phenotype association. Other studies may include more general populations, and their results may differ considerably from those in extreme phenotypes. Variable gene penetrance and expression in dominant conditions can lead to problems in how phenotypes are identified and classified. There may be also variation in diagnosis of different clinical outcomes.

Increased precision can be often achieved by using "intermediate phenotypes," essentially protein markers, from molecular studies that allow phenotypes to be better classified. Such outcomes are also prone to error in measurement.

Design and conduct

Several study designs are used to investigate gene-disease associations, including case-control, cohort, cross-sectional, and family-based designs. There is evidence from other areas of epidemiology that different study designs can yield systematically different estimates of the same underlying association.[6] Furthermore, the conduct of the study (often termed "quality") and the course of events during the study (such as attrition in prospective studies) may be associated with biased estimates. In this review, we address two of the sources of bias with particular implications for genetic association studies: selection of participants and measurement of exposure.

Choice of participants

Systematic differences between cases and controls in a case-control study are an important cause of biased results. Controls should provide an estimate of the exposure distribution from which cases occur, so their selection is crucial to the success of the study. In practice, many studies use convenience samples rather than population-based controls, and so the controls are not derived from the true source population. Poor selection may introduce numerous important differences, including ethnic differences and differences in environmental exposures. These can lead to nonrepresentative genotype distributions and different gene-disease associations.

Prospective cohort studies are generally less susceptible to bias due to choice of participants. However, retrospective cohort and cross-sectional studies run the risk of being inappropriately selective in the participants they choose.

Case-control studies using prevalent cases (for example, from a disease register) may find associations with genetic variants that relate not only to etiology but also to survival (prognostic factors or treatment susceptibility). Studies using incident cases to investigate etiology may not necessarily be affected by this problem. While obtaining truly incident cases in population-based case-control studies is difficult, it is important to assess what proportion of truly eligible cases were included and the reasons why some were not. The etiology of BCHE K variant in Alzheimer disease is a recent example. The association was significant in studies where the age of cases was the age of onset, whereas it was nonsignificant in studies where the recorded age was the age at death.[15]

Definition of the genetic exposure

A number of methods can be used to determine genotype. Genotyping errors can occur because of specific sequence differences in the region of the gene variant being investigated. For example, the assay for Ile462Val polymorphism in CYP1A1 may be interfered with the Thr461Asp polymorphism, depending on the method used. If this "interfering" SNP varies between populations, the error will be both method and population specific. This will usually cause deviation from HWE. Genetic testing quality assurance is an important aspect of study quality.[32]

A similar problem related to misclassification may be observed when some studies use the genotype to classify the observations, whereas some other studies use protein expression. We highlight this problem using an example from meta-analysis of studies relating bladder cancer risk with polymorphisms in *N*-acetyltransferase gene *NAT2*.[34] The genotype is classified as rapid or slow according to the activity of the NAT protein. It has been assessed that genotypes missing two copies of the wild-type allele NAT2*4 are classified as slow acetylating phenotypes. In some bladder cancer case control studies, individuals are classified as rapid or slow according to genotype and according to protein phenotype in others. Putting together these studies in a meta-analysis may cause bias because of misclassification error. The link between genotype and its protein expression is not fully understood and classification of intermediate states can cause problems.

META-ANALYSIS OF GENETIC ASSOCIATION STUDIES

The previous section outlined several sources of genuine variation in gene-disease associations when they are investigated in multiple studies. These clearly have implications for the combination of results across studies in a meta-analysis. Here we discuss some of these implications, along with those

arising from the previous discussion of chance and reporting bias, and some other issues in the statistical synthesis of gene-disease association studies.

Investigating heterogeneity

A substantial literature exists on general approaches to identifying and addressing heterogeneity in meta-analysis.[35–38] Statistical tests for heterogeneity, quantification of inconsistency, random effects models to incorporate variation in true effects, and meta-regression to examine potential reasons for this variation are in common practice. Lohmueller et al.[11] proposed that heterogeneity should be carefully examined by either clinical or statistical methods, and that the magnitude of the heterogeneity should be quantified, wherever possible. Colhoun et al.[4] propose that the main reasons for nonreplication are "all rectifiable," suggesting that in the future, heterogeneity may be much less of a problem. A limitation of virtually all analyses of heterogeneity, particularly investigations of its cause, is that many studies are required before meaningful results are obtained.

Choice of the statistical model

A typical genetic association study categorizes participants into three exposure groups according to genotype (e.g., AA, aA, and aa). Several options are available for the analysis of such data. First, an analysis by alleles (i.e., A vs. a) reduces the data to a 2×2 table in which each participant is represented twice. This approach is not recommended by some researchers because the resulting odds ratio does not reflect the genotype risk and requires an assumption of Hardy-Weinberg equilibrium.[31,39] However, determining equivalence of disease risk between A and a alleles provides strong evidence of a lack of association. Second, a specific mode of inheritance may be assumed from among the dominant, recessive, or codominant (per-allele) genetic models, again reducing the data to a 2×2 table. Attia et al.[39] suggest that the choice of model should express genotype risk in a meaningful way according to some biological background. It is unclear what might constitute such biological background. In many reports of individual association studies, the model applied suggests an underlying assumption about the inheritance mode without being made explicit. A third option is to analyze the data using all three inheritance models, thus performing multiple pairwise comparisons. This creates the risk of the most exciting result being reported, and correction may be appropriate for the multiple comparisons. A final approach is to determine which genetic model is most compatible with the data, by including a parameter to be estimated that distinguishes between the three modes, and possibly intermediate situations.

A common problem for the meta-analyst is how to deal with results reported in different ways. For example, if some studies report only results for carriers and noncarriers (i.e., assuming a dominant genetic model), then other genetic models cannot be applied using standard techniques.

Assessing and addressing Hardy-Weinberg equilibrium

There is some debate over whether individual studies for which the Hardy-Weinberg equilibrium (HWE) assumption does not hold, has not been assessed, or cannot be assessed from the published information should be included in a meta-analysis. Although any departure from HWE should be investigated, many investigators do not believe that departure should be a major criterion for study inclusion/rejection. In any case, it is recommended that a sensitivity analysis be performed with and without these studies, before presenting the results.[39] Especially when applying the per allele model to the analysis, this issue should be carefully investigated because a failure to conform to HWE may indicate that the alleles do not segregate randomly.

Multiple markers

It is common for a meta-analysis, or collection of meta-analyses, to address several polymorphisms on the same gene. The extent to which it is appropriate to combine multiple polymorphisms in the same analysis is undecided. Bellivier et al.[40] advocate a restrictive approach, suggesting that only studies that refer to the same genetic marker, where the effect has been analyzed with the same model, and where the populations were selected according to the same criteria, should be grouped together. Taking a broader perspective, one might argue that every (functional) variant of a gene could be combined in order to maximize power to detect an association between the function of the gene and the risk of disease. An intermediate position would be to combine studies of polymorphisms that have been demonstrated to be in strong linkage disequilibrium.

Problems associated with multiple markers arise in other situations, including the investigation of quantitative traits, where many genetic polymorphisms may contribute small amounts to variability in phenotypes. Specific methods for combining multiple, possibly linked, polymorphisms have not to our knowledge been developed.

Mendelian randomization and covariates

The idea of Mendelian randomization, as we stated in the introduction, gives genetic association studies some notional similarity to randomized trials. But is it sufficient in practice, or should adjustment for covariates be considered? Some authors argue that the results can be confounded due to the distribution of important biomarkers, such as sex and age, both between and within studies. The patterns of linkage disequilibrium in human populations are not well understood: LD can occur at great distances from alleles of interest and can vary substantially between populations.

This is largely a theoretical issue until supportive empirical data become available. Some empirical work on this topic has been undertaken by Taioli and Bonassi[41] who studied the role of biomarkers in the pooled analysis of individual participant data. They concluded that there has been little attention on measuring confounding factors and genetic association studies

are rarely conducted according to the established guidelines for epidemiological studies.

Assessing risk of bias in individual studies

How risk of bias should be measured, and how, or if, these measurements should be incorporated into a meta-analysis is one of the controversial areas of meta-analysis in general. Many now recommend a component-wise approach,[42] whereby important aspects of quality are assessed in a simple manner, but are not combined into a single "quality score." We align ourselves with this approach. No agreed list of important components is widely cited in the literature. In addition to the classical evaluation criteria (such as explicit description of the cases), genetic studies should also be evaluated according to their specific character, such as testing for HWE and control of population stratification.

One of the few resources is the checklist of Little et al.[32] for reporting and appraising genetic studies. They propose that the most important issues to be evaluated are subject selection (selection bias), validity of the genotyping, analysis for population stratification, gene-environment and gene-gene interactions, and statistical power.

An empirical evaluation of the quality of genetic association studies has been conducted.[43] The authors assessed the quality of 40 case-control and cohort studies according to seven criteria: reproducibility of the genotyping method, study blinding, appropriate delineation of cases and controls, adequacy of the cases spectrum, adequacy of controls, and appropriate quantitative analysis. The results showed that more than the half of the studies failed to comply with two or more of these quality requirements.

Publication bias and selective reporting bias

Colhoun et al.[4] propose that the failure to exclude chance and publication bias presents an important problem in replication of findings from gene-disease association studies. In common with issues around heterogeneity, the problem of publication bias has received considerable attention in the general meta-analysis literature.[44] Many of these methods lack power when there are few studies. A recent report has revealed disturbing degrees of selective reporting of outcomes in clinical trials and similar problems doubtless permeate the genetic epidemiology literature.[45] Although no methodological framework has been developed so far to address this problem in its entirety, some developments have been made in the field of clinical trials.[46,47] Such methods could be extended to genetic association studies, although the assumptions required for them may prove complicated and controversial.

Methods to account for multiple tests may have a role in the meta-analysis context.[4] False-positive findings should where possible be identified. Campbell and Rudan[2] suggest an approach for doing this, although a unique definition has not been agreed.

As discussed in the previous section, some interest has focused on empirical comparison of initial findings with later findings.[11,16] The authors have proposed that the result of the first study may represent a false positive and thus should be excluded from the meta-analysis. The resulting drop in power may have important implications for small collections of studies, and so an assessment of the extent of the discrepancy is also pertinent. An investigation of the evolution of results over time, for example using meta-regression, may also indicate early spurious associations. An alternative approach is to perform a "winner's curse" analysis.[11] This method aims to correct for inflations in the *P*-value of the first published study (usually highly significant) by dividing it by a correction factor. This correction factor is the probability of observing an odds ratio at least as large as the one reported in the initial positive study, assuming that the real genetic effect is accurately estimated by a subset of the available studies. This subset is defined by the studies satisfying one of the three characteristics usually met by the first-published study: $P < 0.01$, OR > 2, or use of family-based controls. The empirical evidence also suggests that disagreements between the results of large versus small studies should generally be investigated in a meta-analysis.

Reporting biases would be substantially reduced if studies and findings were reported in full. There will probably always remain unpublished studies of relevance. Whenever available, unpublished information should first be assessed for differences in baseline characteristics compared with the published studies, including biomarkers, allele frequencies, and Hardy-Weinberg equilibrium.[41]

CONCLUSION

Obstacles and opportunities exist in the meta-analysis of genetic association studies. The obstacles include the risk of false-positive findings (Type I error) with associated reporting and publication bias, and numerous sources of true variability within and between populations, including population stratification, epistatic and environmental interactions, and variability of LD within the genome. The opportunities are offered by the enhancement of power (reducing Type II error), the ability to place each study in the context of all others, particularly early spurious results, and the possibility of examining why studies reach different conclusions. We have highlighted the implications for meta-analysis of a number of important statistical issues, including the first study effect, choice of statistical models, and assessing conformity to HWE. A clear understanding of these issues is required before embarking on a meta-analysis in this field.

Meta-analysis of individual patient data from association studies may help overcome many of these problems,[33,39] particularly those associated with incomplete reporting. Standardized definitions can be developed, adjustment for confounding can be performed, alternative genetic models and the role of multiple genes can be assessed and subgroups treated consistently. The collaboration of multiple primary research groups brings further benefits, including the possibility of prospective genotyping of further polymorphisms. However, meta-analyses using individual participant data remain ambitious tasks, cannot eliminate all biases and the full data are not always available retrospectively from investigators.[33]

There is therefore a need for complete and consistent up-front reporting and publishing of genetic association studies to facilitate future meta-analysis. More detailed guidelines are required for the design, analysis, reporting, presentation, and availability of results. The CONSORT statement may provide a useful model for developing similar standards for genetic association studies.[48] Furthermore, depositories for detailed results of all findings, including all null findings, would provide key resources for meta-analyses in this field.

Meta-analyses of genetic association studies, when properly applied and interpreted, contribute to a greater understanding of common, complex diseases. They should both complement and inform the growing collection of large-scale cohort studies of genetic predispositions such as the UK Biobank.[49] The Human Genome Epidemiology Network (HuGE Net)[50] is coordinating and publishing systematic reviews of genetic association studies, an endeavor comparable with the Cochrane Collaboration's database of systematic reviews of the effects of health care interventions.[51] Extensive guidelines for conducting systematic reviews of clinical trials are available,[52] but similar resources for genetic association studies are more limited.[32] We hope that the HuGE Net movement, with others, will continue developing methodology and guidance, so that common standards can be agreed and a growing number of sound meta-analyses be produced to help us separate the wheat from the chaff.

References

1. Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;2:91–99.
2. Campbell H, Rudan I. Interpretation of genetic association studies in complex disease. *Pharmacogenomics J* 2002;2:349–360.
3. Gambaro G, Anglani F, D'Angelo A. Association studies of genetic polymorphisms and complex disease. *Lancet* 2000;355:308–311.
4. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;361:865–872.
5. Whitehead A. Meta-analysis of controlled clinical trials. Chichester: John Wiley & Sons Ltd; 2003.
6. Egger M, Davey Smith G, Altman DG. Systematic Reviews in Health Care; Meta-analysis in context, 2nd ed. London: BMJ Books, 2001.
7. Munafo RM, Flint J. Meta-analysis of genetic association studies. *Trends Genet* In press.
8. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for Meta-analysis in Medical Research. Chichester: John Wiley and Sons Ltd, 2000.
9. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1–22.
10. Little J, Khoury MJ. Mendelian randomisation: a new spin or real progress? *Lancet* 2003;362:930–931.
11. Lohmueller KE, Pearce CL, Pike M et al. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;33:177–182.
12. Ioannidis JP. Genetic associations: false or true? *Trends Mol Med* 2003;9:135–138.
13. Hirschhorn JN, Lohmueller K, Byrne E et al. A comprehensive review of genetic association studies. *Genet Med* 2002;4:45–61.
14. Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000; 405:847–856.
15. Abiola O, Angel JM, Avner P et al. The nature and identification of quantitative trait loci: a community's view. *Nat Rev Genet* 2003;4:911–916.
16. Ioannidis JP, Ntzani EE, Trikalinos TA et al. Replication validity of genetic association studies. *Nat Genet* 2001;29:306–309.
17. Ioannidis JP, Trikalinos TA, Ntzani EE et al. Genetic associations in large versus small studies: an empirical assessment. *Lancet* 2003;361:567–571.
18. Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG et al. Establishment of genetic associations for complex diseases is independent of early study findings. *Eur J Hum Genet* 2004;12:762–769.
19. Eaves IA, Barber RA, Merriman TR. Comparison of linkage disequilibrium in populations from the UK and Finland. *Am J Hum Gen* 1998;A1221.
20. Majewski J, Ott J. GT repeats are associated with recombination on human chromosome 22. *Genome Res* 2000;10:1108–1114.
21. Heath KE, Gahan M, Whittall RA et al. Low-density lipoprotein receptor gene (LDLR) world-wide website in familial hypercholesterolaemia: Update, new features and mutation analysis. *Atherosclerosis* 2001;154:243–246.
22. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003;361:598–604.
23. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* 2000;92:1151–1158.
24. Edland SD, Slager S, Farrer M. Genetic association studies in Alzheimer's disease research: challenges and opportunities. *Stat Med* 2004;23:169–178.
25. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999;65:220–228.
26. Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. *Am J Epidemiol* 1999;149:693–705.
27. Khoury MJ. Fundamentals of genetic epidemiology. In: Kelsey J, ed. Monographs in epidemiology and biostatistics. New York; 1993.
28. Hosking L, Lumsden S, Lewis K et al. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *Eur J Hum Genet* 2004;12:395–399.
29. Gomes I, Collins A, Lonjou C et al. Hardy-Weinberg quality control. *Ann Hum Genet* 1999;63:535–538.
30. Little J, Khoury MJ. Mendelian randomisation: a new spin or real progress? *Lancet* 2003;362:930–931.
31. Little J, Khoury MJ, Bradley L et al. The human genome project is complete. How do we develop a handle for the pump? *Am J Epidemiol* 2003;157:667–673.
32. Little J, Bradley L, Bray MS et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002;156:300–310.
33. Ioannidis JP, Rosenberg PS, Goedert JJ et al. Commentary: meta-analysis of individual participants' data in genetic epidemiology. *Am J Epidemiol* 2002;156:204–210.
34. Hein DW. Molecular genetics and function of NAT1 and NAT2: Role in aromatic amine metabolism and carcinogenesis. *Mutat Res* 2002;506–507:65–77.
35. Higgins JPT, Thompson SG, Deeks JJ et al. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy* 2002;7:51–61.
36. Song F, Sheldon TA, Sutton AJ et al. Methods for exploring heterogeneity in meta-analysis. *Eval Health Prof* 2001;24:126–151.
37. Higgins JPT, Thompson SG, Deeks JJ et al. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–560.
38. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21:1559–1573.
39. Attia J, Thakkinstian A, D'Este C. Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *J Clin Epidemiol* 2003;56:297–303.
40. Bellivier F, Schurhoff F, Nosten-Bertrand M et al. Methodological problems in meta-analysis of association studies between bipolar affective disorders and the tyrosine hydroxylase gene. *Am J Med Genet* 1998;81:349–352.
41. Taioli E, Bonassi S. Methodological issues in pooled analysis of biomarker studies. *Mutat Res* 2002;512:85–92.
42. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42–46.
43. Bogardus ST, Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: the need for methodological standards. *JAMA* 1999;281:1919–1926.
44. Sutton AJ, Song F, Gilbody SM et al. Modelling publication bias in meta-analysis: a review. *Stat Methods Med Res* 2000;9:421–445.
45. Chan AW, Hrobjartsson A, Haahr MT et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457–2465.
46. Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract* 2002;8:353–359.
47. Hutton JL, Williamson PR. Assessing the potential for bias in meta-analysis due to outcome variable selesction within studies. *J R Statist Soc C* 2000;49:359–370.
48. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Clin Oral Investig* 2003;7:2–7.
49. The UK Biobank. Available at: http://www.ukbiobank.ac.uk/ Accessed September 14, 2004.
50. Khoury MJ, Little J. Human genome epidemiologic reviews: the beginning of something HuGE. *Am J Epidemiol* 2000;151:2–3.
51. The Cochrane Library. Chichester: John Wiley & Sons Ltd; 2004.
52. Cochrane Reviewers' Handbook 4.2.2 [updated December 2003] In: Alderson P, Green S, Higgins JPT, eds. The Cochrane Library, Issue 1. Chichester: John Wiley & Sons Ltd; 2004.